

# Comparação de anotações AMR entre línguas: problemas e soluções

Jorge Baptista<sup>1</sup>, Sónia Reis<sup>1</sup>, Pedro A. Santos<sup>2</sup>, João Dias<sup>2,3</sup>

<sup>1</sup> Universidade do Algarve – Faculdade de Ciências Humanas e Sociais / INESC-ID Lisboa – Human Language Technology Lab

<sup>2</sup> Universidade de Lisboa – Instituto Superior Técnico

<sup>3</sup> CISCA – Universidade do Algarve

## Resumo

Este estudo apresenta uma adaptação do esquema de Representação Abstrata de Significado (em inglês, *Abstract Meaning Representation* – AMR) para o português europeu. Esta adaptação, a que se chamou Representação Lexicalizada de Significado (*Lexicalized Meaning Representation* – LMR) foi considerada necessária para lidar com desafios específicos apresentados pela gramática da língua, mas também por várias questões linguísticas suscitadas pela versão atual das diretrizes de anotação de AMR. Alguns destes aspetos foram uma consequência do uso de notação semelhante ao AMR para representar textos reais, do domínio jurídico, que permitisse o seu uso em aplicações de processamento de linguagem natural (PLN). Neste contexto, vários aspetos do AMR foram drasticamente simplificados, por exemplo, a representação de expressões multipalavra, entidades mencionadas e expressões temporais; enquanto outros foram introduzidos *de novo*, tendo sido feitos esforços para manter o esquema de representação o mais compatível possível com a notação em AMR padrão.

**Palavras-chave:** Anotação semântica, representação abstrata de significado (AMR), representação lexicalizada de significado (LMR), processamento da linguagem natural, português europeu.

## Abstract

This study presents an adaptation of the *Abstract Meaning Representation* (AMR) framework for European Portuguese. This adaptation, referred to as *Lexicalized Meaning Representation* (LMR), was deemed necessary to address specific challenges posed by the grammar of the language, as well as various linguistic issues raised by the current version of AMR annotation guidelines. Some of these aspects stemmed from the use of a notation similar to AMR to represent real texts from the legal domain, enabling its use in natural language processing (NLP) applications. In this context, several aspects of AMR were significantly simplified, e.g., the representation of multi-word expressions, named entities, and temporal expressions; while others were reintroduced, with efforts made to maintain the representation scheme as compatible as possible with standard AMR notation.

**Keywords:** Semantic annotation, abstract meaning representation (AMR), lexicalized meaning representation (LMR), natural language processing, European Portuguese.

## 1. Introdução

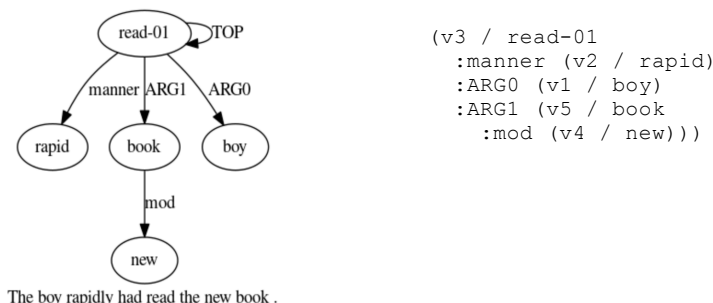
Este estudo pretende contribuir para o desenvolvimento de um quadro teórico e formal para a anotação semântica adequada de textos em língua natural, possibilitando a criação de ferramentas para o seu processamento computacional. A anotação semântica de textos em língua natural visa obter uma representação do significado que seja útil para o desenvolvimento de várias ferramentas e aplicações

(Damonte *et al.*, 2017; Damonte & Cohen, 2018; Seno *et al.*, 2022), particularmente no processamento da linguagem natural (PLN), tais como a desambiguação automática de sentido, a tradução automática, a sumarização de textos ou a geração de documentos multilíngues.

Já foram desenvolvidas várias iniciativas para este propósito. A *Universal Networking Language* (UNL)<sup>1</sup> disponibilizou uma versão da obra *O Príncipezinho*, de Antoine de Saint-Exupéry, com o objetivo explícito de comparar representações do mesmo texto em diferentes línguas. Mais recentemente, a *Abstract Meaning Representation* (AMR) (Banarescu *et al.*, 2013) ganhou popularidade na comunidade de PLN. Este modelo, originalmente proposto para o inglês, visa representar o significado de frases de forma simplificada. Nomeadamente, o significado de cada frase é representado sob a forma de um grafo dirigido acíclico sem raiz, em que os nós são os predicados semânticos (operadores) e os seus argumentos, enquanto os arcos representam as relações semânticas entre os elementos da frase, nomeadamente os papéis semânticos, tal como definidos nas *OntoNotes* (Weischedel *et al.*, 2013), associados aos argumentos dos predicados. Estas últimas constituem uma ontologia que funciona como um ‘catálogo’ de sentidos, servindo de referência aos diferentes sentidos dos elementos predicativos representados no grafo. Outras relações semânticas são igualmente expressas por arcos etiquetados, ligando os predicados a diferentes tipos de elementos e de circunstanciais, substituindo por vezes elementos textuais que veiculam essas relações. Os elementos gramaticais, como os verbos auxiliares, copulativos ou os verbos-suporte são simplesmente suprimidos. Muitos elementos lexicais são substituídos ora por verbos constantes nas *OntoNotes*, ora por outros elementos (e.g. advérbios terminados em *-ly*, substituídos pelos adjetivos morfologicamente associados e ligados a um operador pelo arco etiquetado :MANNER).

A Figura 1 ilustra a representação AMR standard de uma frase simples em inglês,<sup>2</sup> produzida pelo analisador AMREager (Damonte *et al.*, 2017)<sup>3</sup>: *The boy rapidly had read the new book.* ‘O rapaz rapidamente tinha lido o novo livro.’

Figura 1. Representação de uma frase simples em AMR, em formato de grafo (à esquerda) e no formalismo PENMAN (à direita).



<sup>1</sup> [https://en.wikipedia.org/wiki/Universal\\_Networking\\_Language](https://en.wikipedia.org/wiki/Universal_Networking_Language) [data de acesso: 2024/07/19. Todos os URL foram confirmados nesta data.]

<sup>2</sup> Repare-se que, em vez de um nó raiz, este grafo apresenta um arco :TOP que faz um loop sobre o nó que representa o predicado de topo do grafo (read-01); o advérbio *rapidly* ‘rapidamente’ é convertido no adjetivo *rapid* (rápido), que se liga pelo arco :MANNER ao verbo; *boy* (rapaz) e *book* (livro) ligam-se, respetivamente, pelos arcos :ARG0 e :ARG1 ao verbo, permitindo associar a estes argumentos os papéis semânticos definidos para este sentido ‘01’ do verbo *read* (ler), nomeadamente, os papéis de **agente** e de **objeto**; finalmente, repare-se que o auxiliar *had* (*tinha*) é suprimido. Na representação PENMAN, os nós são representados por variáveis (v1, v2, ... v5) que são instanciados pelos elementos lexicais; enquanto os arcos são representados pelas etiquetas precedidas de ‘:’; os diferentes níveis de parentetização representam as relações hierárquicas entre os elementos do grafo; o arco :TOP não é representado.

<sup>3</sup> <https://bollin.inf.ed.ac.uk/amreager.html>

O significado da frase é representado mediante um grafo dirigido acíclico sem raiz (à esquerda) e no formalismo PENMAN (Matthiessen & Bateman, 1991), equivalente (à direita).

Embora tenha sido concebido inicialmente para a língua inglesa e explicitamente rejeite a classificação de *interlíngua*, o AMR presta-se naturalmente à comparação de anotações do mesmo texto em várias línguas (Xue *et al.*, 2014). Este esquema de notação foi, justamente, adaptado para a representação de textos em outras línguas – exemplos incluem anotações da mesma obra em inglês, chinês (Li *et al.*, 2016), espanhol (Miguel Abreira, 2017), turco (Azin & Eryiğit, 2019; Oral *et al.*, 2022), vietnamita (Linh & Nguyen, 2019), português do Brasil (Anchieta, 2020), persa (Takhshid *et al.*, 2022).

Por outro lado, a sua versão inicial tinha como objeto a descrição de frases independentes. Ora, recentemente, as diretivas do AMR foram estendidas para o formalismo UMR (Pustejovsky *et al.*, 2019; Wein & Bonn, 2023) de forma a abranger a anotação de sequências de frases ou discursos (O’Gorman *et al.*, 2018). Naturalmente, as diretivas originais foram ocasionalmente revistas e expandidas para incorporar conceitos que inicialmente não foram suficientemente considerados nas propostas originais (Bonial *et al.*, 2018). Tal como mencionado recentemente por Seno e colegas (Seno *et al.*, 2022), seguindo Hovy e Lavid (2010), estas reformulações e extensões procuram “um equilíbrio necessário entre a profundidade da teoria linguística a ser utilizada e a estabilidade do processo de anotação”, o que não impede “[haver], dentro da comunidade interessada nessa representação semântica, críticos a algumas decisões feitas originalmente [...]” (Seno *et al.*, 2022, p. 51).

O principal objetivo deste trabalho é estabelecer um esquema de anotação a que chamamos *Lexicalized Meaning Representation* (LMR), inspirado nas diretivas do AMR, que visa resolver um conjunto de dificuldades e problemas encontrados nas soluções adotadas até agora. Para tal, foram comparadas as anotações disponíveis de parte da obra de Antoine de Saint-Exupéry *O Príncipezinho*, em inglês,<sup>4</sup> espanhol (Miguel Abreira, 2017) e português do Brasil (Anchieta, 2020), bem como a versão por nós anotada em LMR da mesma obra em português europeu (Baptista, 2024b). Partindo de 50 frases com anotações AMR nas três línguas (considerando duas variedades do português: português do Brasil – PB – e português europeu – PE), é realizada uma análise crítica dos fenómenos observados e das soluções de anotação adotadas, com base nas semelhanças e diferenças entre as anotações.

Considerando aspetos específicos do português europeu, bem como outros requisitos da anotação semântica que, a partir de um ponto de vista do quadro teórico do Léxico-Gramática (Gross, 1981, 1996) desenvolvido a partir da Gramática Transformacional de Operadores harrissiana (Harris, 1991), nos parecem fundamentais, são propostas algumas extensões e reformulações do esquema de anotação AMR padrão (Banarescu *et al.*, 2013). A Tabela 1 apresenta, de forma sintética, as principais diferenças entre as anotações AMR e LMR.

O esquema de anotação *Abstract Meaning Representation* (AMR) baseia-se num “catálogo” de sentidos que tem por base as construções verbais presentes nas *OntoNotes* (Weischedel *et al.*, 2013). Esta metodologia permite tanto a reconstituição como a supressão de elementos textuais, como a inserção de pronomes em posições sintáticas lexicalmente não preenchidas ou a substituição de conjunções e preposições pelas relações semânticas que estas veiculam. No entanto, não inclui a análise de verbos auxiliares, tais como verbos copulativos (*Vcop*) e verbos-suporte (*Vsup*) (ou verbos ditos *leves*). Apenas algumas construções com *Vsup* são consideradas, já que a maior parte dos nomes predicativos é assimilada aos predicados verbais correspondentes (por exemplo, *compra* → *comprar*). Faz ainda uma representação complexa das entidades mencionadas (EM), nomeadamente para a representação normalizada de valores temporais e de quantidade.

---

<sup>4</sup> <https://amr.isi.edu/download/amr-bank-struct-v3.0.txt>. Na altura da redação final deste artigo, esta ligação já não se encontrava a funcionar, mas dada a natureza do corpus, deverá ficar disponível em breve noutro endereço.

Tabela 1. Principais diferenças entre anotações AMR e LMR

<b>Abstract Meaning Representation (AMR)</b> (Banarescu <i>et al.</i> , 2013)	<b>Lexicalized Meaning Representation (LMR)</b>
“Catálogo” de sentidos (predicados semânticos) Verbos: <i>OntoNotes</i> , Weischedel, R. <i>et al.</i> (2013) Outras categorias (nomes e adjetivos) são representadas pelos predicados verbais	“Catálogo” de sentidos: Léxico-Gramática do Português - Verbos: <i>ViPER</i> , Baptista (2012, 2013); <i>Dicionário Gramatical de Verbos do Português</i> , Baptista e Mamede, (2020a); - Nomes predicativos: SNIPER, Baptista e Mamede (2020b)
Grafos dirigidos acíclicos sem raiz (arco :TOP em loop sobre o nó do elemento predicativo principal da frase)	Grafos dirigidos acíclicos mas com raiz (ROOT) ROOT está ligada a elemento predicativo principal (:MAIN), e que é o nó a que se ligam elementos com escopo sobre toda a frase
Reconstituição de elementos reduzidos	Não há reconstituição de elementos reduzidos
Representação em grafo é ‘aposta’ ao texto, sem relação direta (alinhamento) com as palavras/expressões do texto	Relação explícita entre texto e a sua representação, tratando as formas do texto como nós do grafo
Substituição dos elementos predicativos no texto por lemas (sobretudo verbos representados nas <i>OntoNotes</i> , mas também os adjetivos associados aos advérbios terminados em <i>-ly</i> ( <i>-mente</i> ))	Manutenção dos elementos predicativos do texto (associação dos lemas e construção em fase de pós-processamento)
Substituição de elementos textuais (sobretudo gramaticais) pelas relações semânticas que expressam (e.g. conjunções, preposições, etc.); no caso dos advérbios terminados em <i>-ly</i> ( <i>-mente</i> ), substituição do sufixo pela relação :MANNER	Manutenção dos elementos textuais, explicitando relações semânticas (e.g., conjunções, preposições, etc.)
Não considera verbos auxiliares, verbos copulativos nem verbos-suporte (ou verbos leves)	Considera todos os tipos de verbos auxiliares: auxiliares verbais -- temporais, modais e aspetuais; auxiliares adjetivais -- verbos copulativos; auxiliares nominais -- verbos-suporte; auxiliares de construções passivas; considera ainda construções com verbo-operador -- verbo-operador causativo, de ligação, agentivo (Baptista, 2005; Gross, 1981, 1998)
Representação de expressões multipalavra (MWE) de complexidade variável; representação sofisticada de entidades mencionadas (EM) e em particular de expressões temporais e de quantificação	Representação (muito) simplificada de expressões multipalavra (MWE), de entidades mencionadas (EM) e de expressões temporais e de quantificação. Identificação de MWE e de EM em fase de pré-processamento e integração como nós no grafo LMR
Representação de relações anafóricas intrafrásicas. Extensão da notação (O’Gorman <i>et al.</i> , 2018) para relações anafóricas transfrásicas através de cadeias de correferência ao nível do texto	Representação de relações anafóricas intrafrásicas apenas entre elementos explícitos no texto; resolução de anáfora como tarefa de pós-processamento (relações anafóricas transfrásicas ainda não consideradas)
Tratamento muito distinto de predicados verbais (representação standard) e adjetivais (:DOMAIN); construções nominais representadas por construções verbais ( <i>se</i> presentes nas <i>OntoNotes</i> )	Representação homóloga de argumentos de predicados verbais, nominais e adjetivais, correspondente à representação standard: PRED (:ARG0, :ARG1, ...)

Em contrapartida, na *Lexicalized Meaning Representation* (LMR), baseamo-nos no *Dicionário Gramatical de Verbos do Português* (Baptista & Mamede, 2020a), construído sobre a base de dados do léxico gramática dos verbos do português europeu (Baptista, 2012, 2013), para “catálogo” de sentidos das construções verbais, bem como no léxico-gramática dos nomes predicativos, o SNIPER (Baptista & Mamede, 2020b), para os predicados nominais.

Este método privilegia ainda uma representação mais próxima do texto, constituindo-se efetivamente como um processo de anotação, já que a representação se faz diretamente sobre o texto e não é meramente aposta às frases.

Adota-se um nó raiz (ROOT) e uma dependência :MAIN a partir desse nó para o elemento predicativo principal da frase, resolvendo um problema técnico já referido por Anchiêta (2020). Tal solução apresenta ainda a vantagem de, no caso dos adverbiais modificadores de toda a frase (Molinier & Levrier, 2000), por exemplo, o advérbio modalizador *certainly*, poderia assim ser colocado a modificar o nó ROOT<sup>5</sup>:

(EN,195) *I shall certainly try to make my portraits as true to life as possible.*  
'Eu tentarei, certamente, fazer os meus retratos o mais realistas possível.'  
ROOT :MOD (c / certainly)...

em vez de modificar qualquer outro elemento predicativo da frase, como o faz o AMR:

(t / try-01 ... :ARG1 (m / make-02 ... :mod (c / certain))

Adota, também, uma representação simplificada quer das expressões multipalavra (nomes compostos e outras expressões), quer das EM (nomes próprios de pessoas, organizações e locais, por exemplo), quer ainda de expressões temporais e de quantificação, remetendo essa tarefa para uma fase de pré-anotação.

Adota, ainda, uma mesma representação para predicados semânticos de diferentes categorias gramaticais, isto é, verbos, nomes e adjetivos. Assim por exemplo, a par da representação da frase verbal<sup>6</sup>:

(PE,348) – *Isto não importa?!*  
ROOT :MAIN (i1 / importa :ARG0 (i2 / isto) :NEG (n / não))

poderíamos ter encontrado as correspondentes construções adjetivais (1) e nominais (2).

- (1) *Isto não é importante?!*  
ROOT :MAIN (i1 / importante :ARG0 (i2 / isto)  
:NEG (n / não) :VAUX (e / é))
- (2) *Isto não tem importância?!*  
ROOT :MAIN (i1 / importância :ARG0 (i2 / isto)  
:NEG (n / não) :VSUP (t / tem))

<sup>5</sup> Neste artigo, e para mais fácil referência, os exemplos retirados de *The Little Prince/O Pequeno Príncipe/O Principezinho* são apresentados com indicação do número da frase na representação AMR da respetiva edição: EN, para edição inglesa; ES, para as frases traduzidas (da edição inglesa) para espanhol (Migueles Abreira, 2017); PB, as frases da tradução (a partir da edição francesa) para português do Brasil (Anchiêta, 2020); e PE, da tradução (a partir da edição francesa) para português de Portugal (Baptista, 2024b). Indica-se depois uma tradução livre do exemplo. Finalmente, apresenta-se a representação da frase no formalismo PENMAN, eventualmente sem indentação e por vezes truncada para maior concisão, deixando apenas os elementos relevantes para a discussão. Já os exemplos construídos para comparação com os exemplos originais são, por seu turno, numerados sequencialmente.

<sup>6</sup> À semelhança do AMR, em LMR a indicação do tipo de frase faz-se com operadores específicos (e.g., :MODE-EXCLAMATIVE), não se marcando as frases declarativas. Nestes exemplos, esta notação do tipo de frase foi omitida, por simplicidade.

Note-se que, em AMR standard, as frases adjetivais e nominais, formalmente idênticas aos exemplos acima, poderiam ser representadas de diversas formas, dependendo, nomeadamente, da existência prévia de uma representação da construção (*frame*) do elemento predicativo na ontologia de significados. Assim, no exemplo (EN,461), recorre-se ao conceito de :DOMAIN.<sup>7</sup> Noutro exemplo (EN,358), ao facto de o elemento predicativo, neste caso um adjetivo, se encontrar subordinado a outro operador já conduz a outra notação:

(EN,461) *That is of no importance*  
 ‘Isso não tem qualquer importância’  
 (i / important-01 :polarity- :ARG1 (t / that))

cp. *That is not importante*  
 ‘Isso não é importante’  
 (t / that :domain (i / important :polarity-))

(EN,348) *You think that is not important !*  
 ‘Tu pensas que isso não é importante’  
 (t / think-01 :ARG0 (y / you)  
 :ARG1 (t2 / that :ARG1-OF (i / important :polarity-))

A proposta de representação do LMR abrange, além disso, diversos tipos de verbos auxiliares:

- *verbais* – temporais, modais e aspetuais;
- *adjetivais* – verbos copulativos;
- *auxiliares nominais* – verbos-suporte;
- *auxiliares de construções passivas* (*ser*, *estar* e suas variantes);

Considera, além destes, as construções com *verbo-operador*, *verbo-operador causativo*, *verbo-operador de ligação*, e *verbo-operador agentivo* (Baptista, 2005; M. Gross, 1981). A fim de clarificar alguns destes conceitos, apresenta-se exemplos destas construções e a sua representação. Assim, a frase seguinte (PE,445), representa o verbo auxiliar modal *dever*, tratando-o como elemento que modifica com esse valor gramatical específico o verbo auxiliado (*regressar*), como se pode ver abaixo:

(PE,445) *Julgava que não devia regressar*  
 (ROOT :MAIN (j / julgava  
 :ARG1 (q / que  
 :OP2 (r / regressar  
 :VAUX (d /devia  
 :NEG (n / não))))))

Em virtude do contexto em que foi inicialmente desenvolvido, nomeadamente, para a análise de texto jurídico, o LMR dá um tratamento especial às construções de verbos auxiliares que exprimem a modalidade *deontica* (*dever* + infinitivo, *ter de* + infinitivo), recorrendo a um operador específico, :DEVER. No exemplo, porém, *dever* tem uma modalidade *epistémica*, como se pode verificar pelo original em inglês:

(EN,445) *He believed that he would never want to return*

<sup>7</sup> <http://www.isi.edu/~ulf/amr/lib/popup/domain.html>

‘Ele acreditava que nunca haveria de regressar’,

em que se emprega o auxiliar *would*. Por essa razão, foi usado o operador genérico para auxiliares :VAUX.

Por seu turno, os verbos-suporte (*Vsup*) são elementos auxiliares de nomes predicativos. Para a sua representação, é instanciada uma relação :VSUP que liga o nome predicativo ao *Vsup* com que este se constrói. Assim, na frase seguinte (PE,196), o verbo *ter* é considerado o suporte do nome predicativo *certeza* e é, pois, representado como um elemento modificador deste nome. Note-se a homologia entre o tratamento dos auxiliares e os respetivos predicados que eles auxiliam:

```
(PE,196) Mas não tenho a certeza de conseguir
(ROOT :MOD (m / mas)
 :MAIN (c1 / certeza
 :VSUP (t / tenho)
 :NEG (n / não)
 :ARG1 (c2 / conseguir)))
```

Note-se, ainda, que o verbo *conseguir* é, por sua vez, um auxiliar de um predicado elidido (*fazer retratos*) e que se refere ao conteúdo da frase anterior:

```
(PE,195) Vou tentar, obviamente, fazer retratos o mais fiéis possível.
```

Contudo, como em LMR não se faz reconstituição de elementos anafóricos que não estejam presentes *na mesma frase*, o verbo *conseguir* é deixado apenas como ARG1 de *certeza*.

Estas adaptações do AMR para o LMR são explicitamente codificadas nas diretivas de anotação especificamente desenvolvidas para o efeito (Baptista, 2024a).<sup>8</sup> Os demais aspetos apresentados na Tabela 1 serão abordados *in situ*, no contexto da análise das frases selecionadas para as secções subsequentes.

Estas diferenças metodológicas entre o AMR e o LMR resultam de abordagens em parte distintas na representação semântica de textos: embora cada uma apresente as suas vantagens e desafios específicos, o LMR distingue-se por procurar conciliar a precisão da representação semântica com a fidelidade ao texto que lhe está na base, sugerindo uma abordagem potencialmente mais eficiente na análise semântica. Efetivamente, algumas das inconsistências observadas na representação disponível de frases em AMR poderá, em parte, resultar justamente nos diferentes princípios adotados.

Este trabalho foi desenvolvido nas seguintes etapas:

- *Alinhamento* das frases nas diferentes línguas/variedades, considerando as especificidades de cada língua e a tradução respetiva a partir da edição original da obra (a edição em inglês, no caso da versão espanhola; a edição francesa, no caso das traduções portuguesas) e resolvendo os problemas de alinhamento encontrados.
- *Anotação* das frases na versão em português europeu de *O Príncipezinho*, realizada independentemente por dois anotadores e com base num conjunto de diretivas (Baptista, 2024a).<sup>9</sup> desenvolvido autonomamente. Estas diretivas procuram (a) adaptar o AMR a situações linguísticas observadas no português, mas não observadas no inglês; (b) tornar sistematicamente explícita e consistente a relação entre o texto e a anotação; e (c) dar conta de forma adequada de fenómenos linguísticos relevantes não contemplados pelo quadro do AMR. Após esta fase, as anotações foram

<sup>8</sup> [https://gitlab.hlt.inesc-id.pt/u000803/lmr4pt/-/blob/master/LMR\\_Directives\\_20240501.pdf](https://gitlab.hlt.inesc-id.pt/u000803/lmr4pt/-/blob/master/LMR_Directives_20240501.pdf)

<sup>9</sup> [https://gitlab.hlt.inesc-id.pt/u000803/lmr4pt/-/blob/master/public/LMR4PT\\_Principezinho.pdf](https://gitlab.hlt.inesc-id.pt/u000803/lmr4pt/-/blob/master/public/LMR4PT_Principezinho.pdf)

comparadas e discutidas, de forma a se chegar a uma forma de representação final, única e consensual (Baptista, 2024b).

- Finalmente, e para este artigo, foi realizada uma *comparação* sistemática e reflexão crítica das anotações das frases nas diferentes línguas.

Este trabalho está organizado do seguinte modo: Após esta breve apresentação do conjunto de diretivas da LMR desenvolvidas para a anotação de textos em português europeu; apresenta-se um conjunto de frases em língua natural (em quatro línguas/variedades), alinhadas manualmente e comparadas de forma sistemática, com menção explícita dos fenómenos linguísticos encontrados e das soluções de representação adotadas (ou falta delas); e, por último, faz-se uma análise crítica das limitações observadas nas anotações disponíveis para as frases estudadas.

## 2. Comparação interlinguística das anotações AMR

Nesta secção, comparamos a representação AMR de *O Príncipezinho* em três línguas (inglês, – EN –, espanhol – ES – e português – PT), considerando duas variedades de português: português europeu (PE) e português do Brasil (PB). Como mencionado anteriormente, apenas o conjunto de 50 frases em espanhol (ES) foi considerado para esta tarefa de comparação, já que só para estas frases do romance é possível fazer uma comparação entre as várias línguas/variedades, não havendo, tanto quanto sabemos, outra anotação em ES.

Alguns aspetos devem ser considerados antes da análise propriamente dita. Primeiro, o material fonte para as traduções. Como sabemos, o romance foi originalmente escrito em francês, mas publicado simultaneamente, em 1943, tanto em francês como em inglês. A estrutura genológica do texto não é a nossa preocupação aqui. No entanto, notamos que a versão em ES foi traduzida diretamente da versão em EN pelo autor do *corpus* AMR. Por outro lado, os textos PB e PE foram anotados em AMR com base em edições existentes do romance, tendo, aparentemente, sido usada a edição em francês como base para essas traduções.

Um segundo aspeto a considerar: a liberdade dos tradutores. Naturalmente, os tradutores têm um certo grau de liberdade ao adaptar o material fonte. Mesmo no caso das traduções FR>PB e FR>PE, em cada uma se encontra diferentes escolhas (embora frequentemente equivalentes). Por exemplo, na frase com id=43:

(PB,43) *Imagem então a minha surpresa, quando, ao despertar do dia, uma vizinha estranha me acordou.*

(PE,43) *Assim, imaginem a minha surpresa quando, ao amanhecer, fui despertado por uma vizinha muito curiosa.*

um dos tradutores usou o advérbio conjuntivo *então* (PB) enquanto o outro escolheu o advérbio *assim* (PE). No material fonte em FR:

FR *Alors vous imaginez ma surprise, au lever du jour, quand une drôle de petite voix m'a réveillée.*

‘Então, imaginem a minha surpresa, ao amanhecer, quando uma vizinha estranha me acordou.’

encontra-se o advérbio equivalente *alors* ‘então/assim’. Repare-se, também, na escolha lexical dos adjetivos *estranha* (PB) e *curiosa* (PE) para a expressão FR *une drôle de petite voix* ‘uma vizinha estranha’. Finalmente, repare-se o uso do diminutivo *vizinha* para a expressão original *petite voix*. Diferenças lexicais como estas, desde que não afetem a representação AMR, não são consideradas aqui. Pelo contrário, escolhas lexicais, tais como *ao despertar do dia* (PB)/*ao amanhecer* (PE), *me acordou* (PB)/*fui despertado* (PE), se



tiverem sido diferentemente representadas, podem ter impacto na comparação, pelo que foram consideradas aqui.

As diferenças estendem-se também à pontuação. Em PE, o advérbio foi anteposto e separado da frase por vírgula. Em PB, o advérbio, embora sintaticamente separado, como se pode ver pela prosódia, não é separado por vírgulas. Na versão FR consultada, o advérbio *alors* também é anteposto, mas não é separado por vírgula. Diferenças de pontuação como estas, desde que não afetem a representação AMR, não são consideradas aqui.

Finalmente, verifica-se pontualmente a omissão de elementos lexicais ou expressões na representação AMR. É o que sucede na tradução da frase acima em ES, em cuja representação AMR o advérbio *por tanto* foi omitido, sendo substituído por um operador *causa* no nó de topo da frase:

(ES,1[=43]) *Por tanto, puedes imaginarte mi asombro cuando una extraña vocecita me despertó al amanecer.*  
 ‘Portanto, podes imaginar o meu espanto quando uma vozinha estranha me acordou ao amanhecer.’

Figura 2. ES; frase id1[=43]

```
(c / causar
  :ARG1 (p / posible
    :ARG1 (i / imaginar
      :ARG0 (t / tú)
      :ARG1 (a / asombrar
        :ARG1 (y / yo)
        :tiempo (a2 / amanecer
          :tiempo-de (d / despertar
            :ARG1 y
            :ARG0 (v / voz
              :mod (e / extraño)
              :mod (p2 / pequeño))))))))
```

Efetivamente, tanto na versão FR como na versão EN originais, observa-se um advérbio: *alors/thus*:

(EN,43) *Thus you can imagine my amazement , at sunrise , when I was awakened by an odd little voice.*  
 ‘Assim, podes imaginar o meu espanto, ao nascer do sol, quando fui acordado por uma vozinha estranha.’

Figura 3. EN; frase id=43

```
(c / cause-01
  :ARG1 (p / possible-01
    :ARG1 (i2 / imagine-01
      :ARG0 (y / you)
      :ARG1 (a / amaze-01
        :ARG1 (i / i)
        :time (s / sunrise
          :time-of (w / wake-01
            :ARG0 (v / voice
              :mod (o / odd)
              :mod (l / little))
            :ARG1 i))))))
```

A solução descritiva do AMR em ES é, pois, igual à da representação em EN e praticamente equivalente à solução descritiva adotada na adaptação do AMR para português do Brasil. Ora, na representação em PB (ver Figura 4, abaixo), o advérbio correspondente (*então*) foi simplesmente omitido, perdendo-se, assim, parte do significado que veicula na frase.

Como se disse anteriormente, o esquema de anotação aqui apresentado foi inicialmente desenvolvido num contexto particular, de anotação de textos jurídicos. Por essa razão, e ao contrário das diretivas do AMR, as palavras no texto são mantidas nos grafos LMR produzidos neste contexto e não se encontram para já lematizadas. Tal implica uma etapa de pós-processamento para produzir essa lematização. Esta diferença não é levada em consideração na comparação. Na mesma linha, em LMR, a classe gramatical das palavras no texto não é alterada, ao contrário das diretivas AMR que, no caso de nominalizações, preconiza que se indique sempre o lema verbal (por exemplo, *surpresa* → *surpreender*) em que, em princípio, terá sido descrita a respetiva grelha argumental. Finalmente, no esquema de anotação AMR, os grafos dirigidos acíclicos (DAG) são construídos sem um nó raiz, desenvolvendo-se diretamente a partir do operador superior da frase. Este aspeto foi criticado, entre outros, por Anchiêta *et al.* (2019). Considerando pertinentes tais observações, sobretudo tendo em vista a comparação automática de anotações diferentes da mesma frase, propusemos igualmente para a anotação LMR a inserção de um nó raiz, a partir do qual se desenvolve o grafo que descreve o significado de cada frase. Esta decisão, contudo, não afeta a conversibilidade das notações LMR no esquema do AMR standard.

## 2.1. Aspetos na comparação PB/PE

Começamos por comparar a frase com o id=43. As frases em PB e PE e as respetivas notações AMR são apresentadas nas Figuras 4 e 5. Ao comparar as duas notações desta frase, o aspeto mais saliente é a falta de representação do modo imperativo (*imaginem*) em PB. No AMR padrão, espera-se que o traço :MODE IMPERATIVE tivesse sido incluído no nó do verbo *imaginar*.

(PB,43) *Imaginem então a minha surpresa, quando, ao despertar do dia, uma vizinha estranha me acordou.*

Figura 4. PB; frase id=43

```
(i / imaginar-01
  :ARG1 (s / surpreender-01
    :ARG0 (m / minha)
    :time (d / despertar
      :time-of (a / acordar-01
        :ARG0 (v / vizinha
          :mod (e / estranha))))))
```

(PE,43) *Assim, imaginem a minha surpresa quando, ao amanhecer, fui despertado por uma vizinha muito curiosa.*

Figura 5. PE; frase id=43

```
(ROOT :MODE-IMPERATIVE ()
  :CAUSE (a / assim)
  :MAIN (i / imaginem
    :ARG1 (s / surpresa
      :ARG0-OF (m1 / minha))
    :TIME (q / quando
      :OP2 (d / despertado
        :VAUX (f / fui))
```

```
:ARG0-OF (v / vozinha
           :ARG0-OF (c / curiosa
                    :DEGREE (m2 / muito)))
:TIME (aa / ao_amanhecer)))))
```

O advérbio conjuntivo *então* (Català *et al.*, 2020; Molinier & Levrier, 2000; Muller *et al.*, 2023; Palma, 2009), que funciona como um modificador de toda a frase, ligando-a ao discurso anterior, foi ignorado na representação em PB. Em PE, foi usada uma expressão equivalente, com o advérbio *assim*, que é um advérbio do mesmo tipo sintático. Este tipo de advérbios foi explicitamente considerado nas diretivas de notação AMR adaptadas ao PE. Nestas, preconiza-se que este advérbio esteja ligado ao nó ROOT e o seu papel semântico de **causa** seja explicitamente codificado.

O sujeito de *surpreender/surpresa* é denotado pelo pronome possessivo *minha*. Apesar das diferentes notações ARG0/ARG0-OF, relacionadas com o facto de a notação PE respeitar a classe gramatical das palavras do texto, as duas notações podem ser consideradas equivalentes.

A seguir, repare-se que, em PB, a expressão temporal *ao despertar do dia* corresponde à expressão *ao amanhecer* em PE. O valor idiomático da expressão PB *ao despertar do dia* corresponde à expressão idiomática em PE *ao nascer do dia*, embora este não tenha sido utilizado na tradução. De qualquer forma, ambas as expressões adverbiais temporais modificam *acordar* em PB e *despertar* em PE. Na representação AMR da frase brasileira, a expressão fixa *ao despertar do dia* não é identificada como tal, e *dia* é omitido. Para todos os efeitos, é como se se tratasse do verbo simples *despertar* (e não da expressão idiomática *ao despertar do dia*) ou, talvez, do nome *despertar*.

O verbo *despertar* não aparece associado a nenhuma construção verbal, o que se indica por *não* vir assinalado com um índice numérico -01, não estando sequer representado no catálogo do Verbo-Brasil (Duran & Aluísio, 2015)<sup>10</sup> -- se se considerasse um nome, talvez essa notação estivesse correta. Contudo, a principal imprecisão, quanto a nós, reside no facto de, aparentemente, *despertar* ter sido interpretado como modificador (temporal) de *surpreender* e não de *acordar*, o que nos parece sintaticamente inadequado. Efetivamente, o adverbial *ao despertar do dia* intercalado entre a conjunção subordinativa *quando* e a oração subordinada *uma vozinha estranha me acordou*, não pode ser interpretado como um modificador do verbo da oração principal *surpreender-01* (no texto, este corresponde ao nome *surpresa*).

Quando esta solução é comparada com as frases em inglês e francês, que retomamos abaixo:

(EN,43) *Thus you can imagine my amazement, at sunrise, when I was awakened by an odd little voice.*

‘Portanto, podes imaginar o meu espanto, ao nascer do sol, quando fui acordado por uma vozinha estranha.’

FR: *Alors vous imaginez ma surprise, au lever du jour, quand une drôle de petite voix m’a réveillé.*

‘Então, podem imaginar a minha surpresa, ao amanhecer, quando uma vozinha estranha me acordou.’

verifica-se que os advérbios temporais correspondentes (*at sunrise/au lever du jour*) podem funcionar quer como modificadores de *amazement/surprise* (leitura preferencial), quer como modificador de *awakened/réveillé*. Assim, parece que o PB seguiu a notação EN de forma demasiado próxima:

```
(a / amaze-01 :ARG1 (i / i) :time (s / sunrise ...
```

<sup>10</sup> <http://www.nilc.icmc.usp.br/semanticnlp>

Repare-se, ainda, que em PB, *acordar* está numa construção ativa, enquanto que o verbo correspondente em PE, *despertar*, está numa construção passiva. As diferenças de representação destas frases daí decorrentes não são relevantes para a comparação, já que ambas as construções são consideradas como equivalentes aqui.

Finalmente, tanto a notação AMR de PB quanto a de EN introduzem uma dependência de tempo (REIFICATION?) entre *despertar/sunrise* e os verbos *acordar/wake*. A definição desta dependência nas diretivas é incerta, mas pode ter sido apropriadamente usada, por exemplo, no caso da frase com id=434:

(PB,434) *Na manhã da partida, pôs o planeta em ordem.*

onde *da partida* é formalmente um complemento de *manhã* e este último é um modificador temporal da frase principal. Portanto, a solução descritiva fornecida no AMR:

Figura 6. PB; frase id=434

```
(p / por-01
  :ARG0 (e / ele)
  :ARG1 (p1 / planeta
    :mod (e1 / em-ordem))
  :time (d / date-entity
    :dayperiod (m / manhã)
    :time-of (p2 / partida)))
```

parece relativamente consistente com os outros usos de papéis “invertidos” (assinados pelo sufixo -of) conforme definidos nas diretivas AMR. No entanto, nesta frase com id=43, esta dependência TIME-OF não parece ser consistente.

Em síntese: as principais vantagens da notação LMR em relação à notação AMR padrão consistem em:

- i. indicar explicitamente o nó ROOT e a partir dele marcar o operador ou elemento predicativo principal;
- ii. ser capaz de ligar ao nó ROOT diferentes tipos de elementos predicativos, como advérbios modificadores de frases, eventualmente ligando essa frase a uma frase anterior no discurso, ou associando-lhe o modo/modalidade do tipo de frase, por exemplo, :MODE-IMPERATIVE;
- iii. representar explicitamente os elementos predicativos na sua classe gramatical, baseando-se numa representação léxico-sintática mais sofisticada (um léxico-gramática) subjacente à representação semântica, sem ter de converter outras categorias morfossintáticas em verbos, como faz o AMR;
- iv. representar explicitamente a conjunção subordinativa e os verbos auxiliares, incluindo os das construções passivas; e
- v. simplificar a representação de expressões multipalavra (ou palavras compostas, por exemplo, o advérbio composto *ao amanhecer*).

De seguida, descrevemos a alternância entre frases em discurso direto e indireto, conforme mostrado na Figura 7 e na Figura 8. Primeiro, descrevemos as diferenças não relevantes entre as frases nas duas variedades e, em seguida, comparamos as diferentes soluções de anotação do AMR padrão e do LMR.

(PB,64) *Respondeu-me: – Não tem importância.*

Figura 7. PB; frase id=64

```
(r / responder-01
  :ARG0 (e / ele)
  :ARG1 (i / importância :polarity -
        :mod (t / ter-01))
  :ARG2 (e2 / eu))
```

(PE,64) *Ele respondeu: – Não faz mal.*

Figura 8. PE; frase id=64

```
(ROOT :MAIN (respondeu
  :ARG0 (e / ele)
  :ARG1 (fm / faz_mal
        :NEG (n / não))))
```

As duas frases são diferentes na medida em que a PB omite o sujeito de 3.<sup>a</sup> pessoa (o **agente-locutor**, speaker) mas mantém explícito o complemento indireto de 1.<sup>a</sup> pessoa (o **alocutário**, addressee); enquanto a tradução PE faz precisamente o oposto. Na representação AMR, o sujeito omitido é reconstruído, conforme as diretivas de notação padrão; enquanto o LMR se atém à superfície do texto, representando apenas os elementos explícitos na frase. Embora a reconstrução de um sujeito omitido não seja uma tarefa insuperável, entre outros aspetos devido à concordância verbal (Baptista *et al.*, 2016), a reconstrução de complementos omitidos é muito mais complexa, de um ponto de vista do PLN. A solução LMR implica uma etapa de resolução de anáfora, numa fase de pós-processamento, para recuperar o complemento indireto omitido a partir da informação léxico--sintática associada ao verbo (classe 09I; Baptista & Mamede, 2020a).

Outra diferença é a representação da negação, que, no AMR padrão, é representada por um traço (:polarity-) que aparece ligado ao substantivo predicativo *importância*; enquanto, no LMR, uma dependência :neg é usada para representar explicitamente o advérbio de negação *não*.

Em segundo lugar, embora ambas as frases apresentem o verbo *dicendi* (Baptista, 2010) *responder*, introduzindo uma fala de discurso direto da mesma maneira (incluindo pontuação), o conteúdo dessa fala é diferente: A fala do PB emprega a construção, que analisamos como uma construção de verbo-suporte *ter importância* (Baptista *et al.*, 2022; Baptista & Mamede, 2020b; Baptista & Mamede, 2023); enquanto a versão do PE usa a expressão idiomática verbal *fazer mal* (classe C1PN; Galvão *et al.*, 2019a; Galvão *et al.*, 2019b). As duas notações são, além disso, equivalentes, no que diz respeito à representação da relação entre o verbo *dicendi* e o discurso direto que aquele introduz. Em ambas, o operador principal da frase em discurso direto encaixada sob o verbo *dicendi responder* é analisado como *arg1* desse verbo. Isso é relevante, pois a mesma representação (ou outra muito semelhante) pode ser produzida para as correspondentes frases equivalentes em discurso indireto, como, por exemplo, PB: *Ele respondeu-me que (isso) não tinha importância* / PE: *Ele respondeu-me que (isso) não fazia mal*. No entanto, enquanto em AMR padrão o verbo-suporte *ter* deveria ter sido ignorado<sup>11</sup>, o LMR teria representado explicitamente: :ARG1 (i / importância :VSUP (t / tem)). Em vez disso, no AMR, o verbo-suporte é analisado como um :mod do substantivo predicativo, o que é inconsistente com as diretivas do AMR padrão.

Em resumo, podemos ver que:

- i. a representação da relação entre o verbo *dicendi* e a frase em discurso direto é praticamente a mesma no AMR padrão e no LMR;
- ii. a negação é representada de maneira equivalente, mas o LMR representa explicitamente o advérbio de negação no gráfico; e

<sup>11</sup> [https://www.isi.edu/~ulf/amr/lib/amr-dict.html#light\\_verb\\_construction](https://www.isi.edu/~ulf/amr/lib/amr-dict.html#light_verb_construction)

- iii. apenas o LMR representa o verbo-suporte nas construções de nomes predicativos, parecendo haver aqui inconsistência entre a anotação adotada para a frase em PB e as diretivas de anotação do AMR padrão.

No próximo exemplo (Figura 9 e Figura 10), discutimos a distinção entre complementos argumentais (essenciais) e não argumentais (adjuntos).

(PB,65) *Desenha-me um carneiro.*

Figura 9. PB; frase id=65

```
(d / desenhar-01
:ARG0 (e1 / ele)
:ARG1 (c / carneiro)
:ARG2 (e / eu))
```

(PE,65) *Desenha-me uma ovelha.*

Figura 10. PE; frase id=65

```
(ROOT :MODE-IMPERATIVE
  :MAIN (d / desenha
    :ARG1 (o / ovelha)
    :BENEFICIARY (m / me)))
```

Primeiro, as duas frases são quase idênticas, exceto pela escolha de palavras dos tradutores (por exemplo, PB: *carneiro* / PE: *ovelha*), o que não é relevante aqui. A principal diferença é a forma como a anotação brasileira considera o pronome dativo *-me*, que é analisado como um :ARG2, isto é, como um argumento de *desenhar*-01. Consultando o catálogo de sentidos verbais do Verbo-Brasil (Duran & Aluísio, 2015), que fundamenta a representação AMR, verifica-se que existe apenas um sentido de palavra para este verbo, e que ele é descrito como tendo 3 argumentos (incluindo um ARG2 com um papel semântico de **beneficiário**). No recurso lexical correspondente em português, o *Dicionário Gramatical de Verbos do Português* (Baptista & Mamede, 2020a), usado como referência para a anotação LMR, o verbo (também uma única entrada) é descrito como tendo apenas dois argumentos: um **agente** e um **objeto** (classe 32A). Uma vez que, em português europeu, consideramos este complemento (facultativo) como um constituinte não essencial – um adjunto (Leclère, 1976; ver Van Peteghem (2006), para uma discussão), isto é, um elemento que não pertence ao domínio argumental deste verbo *desenhar*, é adicionada uma dependência :BENEFICIARY, ligando o verbo ao pronome.

Embora se trate de uma frase imperativa em discurso direto, a representação AMR não inclui essa notação (:MODE IMPERATIVE no nó do verbo correspondente), prevista nas respetivas diretivas. Pelo contrário, em LMR a anotação do modo imperativo da frase é feita diretamente no nó ROOT.

Repare-se que o sujeito da frase é reconstituído, em AMR, como um *ele*, embora na realidade se trate de uma 2ª pessoa (*tu/você*). Efetivamente, o verbo está conjugado na 2ª pessoa do imperativo (*desenha*) e não na flexão de uma 2ª pessoa na forma de tratamento por *você*. Esta notação parece, pois, menos adequada ao valor de pessoa do sujeito de uma frase imperativa, já que será igual à de uma mera 3ª pessoa em frases de outro tipo (declarativo). Em contraponto, em LMR, os elementos elididos não se representam nesta fase, razão pela qual não ocorre o :ARG0.

Note-se ainda que esta mesma frase ocorre mais vezes no texto. Tal permite verificar algumas inconsistências na anotação. Por exemplo, na frase (id=74) a representação é exatamente a mesma.

Já na frase (id=46), tudo se representa como nos exemplos anteriores (id=65, 74), exceto o :ARG0, que está omitido, ao contrário do que indicam as diretivas AMR, que preconizam que os argumentos elípticos devam ser reconstituídos sob uma forma pronominal.

Noutra frase ainda,

(id=44), *Dizia: -- Por favor... desenha-me um carneiro!*

em que o verbo *desenhar* se encontra em discurso direto mas encaixado sob um *verbum dicendi* (*dizer*), o verbo *desenhar* aparece representado como :ARG1: do verbo *dizer*; aqui já é indicado o modo imperativo, ainda que o :ARG0 não seja representado. Não se representa aqui, porém, a expressão de polidez *por favor*, presente na frase (ver adiante).

Na frase com o id=60, igualmente com o verbo *desenhar* encaixado como ARG1 sob um *verbum dicendi* (*repetir*); o verbo *repetir*, por sua vez, é representado com um ARG0 correspondente ao **locutor** e um ARG2 correspondente ao **interlocutor**. O verbo *desenhar* mantém a estrutura com 3 argumentos. O seu :ARG0 é representado por um *eu*, correspondente ao **interlocutor** de *repetir*; não se representa a modalidade

imperativa, mas aparece o traço :POLITE +,<sup>12</sup> provavelmente derivado da presença de *por favor*. Note-se que esse traço não estava presente na frase (id=44), ainda que essa expressão lá figurasse.

Ainda sobre *por favor*, além das duas frases indicadas acima (id= 44 e 60), a expressão ocorre ainda duas vezes mais. Na frase com o (id=1114), -- *Por favor... cativa-me! disse ela.*, esta expressão aparece representada como uma locução (ou expressão multipalavra) *por-favor*, e como um modificador de maneira/modo (:MANNER) do verbo *dicendi* (dizer-01). Embora se representem o :ARG0 (*ela*) e o :ARG1 (*cativar*), não se representa o :ARG2 de *dizer*, que corresponderia ao **interlocutor**. Também não é representada a modalidade imperativa, nem no verbo *dizer* nem no verbo *cativar*, apesar de este verbo se encontrar conjugado numa 2ª pessoa do singular (*tu*).

Finalmente, na frase:

(id=1561;1562) *Então, por favor, não me deixem tão triste: escrevam-me depressa que ele voltou...*,

também não há indicação da modalidade imperativa, apenas da negação (:polarity-) no verbo *deixar*, assim como o traço de polidez (:POLITE +). Note-se que o :ARG0 de *deixar* não é representado, apesar de este verbo se encontrar conjugado numa 2ª pessoa do plural (*vocês*).

Parece, pois, haver, alguma inconsistência na anotação, quer relativamente à marcação da modalidade imperativa, quer à indicação do traço de polidez, associado a *por favor*, quer ainda aos argumentos essenciais dos predicados que se encontram elididos, mormente o :ARG0 e o :ARG2.

O caso seguinte (Figura 11 e Figura 12) considera diferentes tipos de complementos locativos e introduz o tema das orações subordinadas relativas.

(PB,72) *Tudo é pequeno onde eu moro.*

Figura 11. PB; frase id=72

```
(p / pequeno
  :mod (t / tudo)
  :location (m / morar-01
    :ARG0 (e / eu)))
```

(PE,72) *Moro num sítio muito pequenino.*

Figura 12. PE; frase id=72

```
(ROOT :MAIN (m1 / moro
  :LOCATION (e / em
    :OP2 (s / sítio
      :ARG0-OF (p / pequenino
        :DEGREE (m2 / muito))))))
```

Começamos por reconhecer as diferentes escolhas dos tradutores, uma vez que as fontes em francês e inglês (Figura 13) também são bastante diferentes:

FR: *Chez moi c'est tout petit.* ‘Em minha casa, é tudo pequeno.’

(EN,72) *Where I live, everything is very small.* ‘Onde eu moro, tudo é muito pequeno.’

<sup>12</sup> <https://www.isi.edu/~ulf/amr/lib/amr-dict.html#:polite>





Figura 13. EN; frase id=72

```
(s / small
  :degree (v / very)
  :domain (e / everything)
  :location (l2 / live-01
             :ARG0 (i / i)))
```

A frase em PE é, de certa forma, mais próxima da estrutura em FR, no sentido de que o locativo é expresso por um complemento preposicional FR: *chez moi*/PE: *num sítio muito pequenino* (poderá ser difícil fornecer uma tradução direta para a preposição francesa *chez*.)

Este tipo de complementos locativos é representado no AMR padrão por uma dependência :LOCATION, dependendo do verbo principal (e a preposição locativa é ‘consumida’ por esta dependência). Ao contrário, a notação LMR, embora usando a mesma dependência, representa explicitamente a preposição. Também nesta frase PE, o núcleo do complemento locativo, o substantivo *sítio*, é modificado pelo adjetivo *pequenino*, que por sua vez é modificado por um advérbio :DEGREE *muito* (o facto de *pequenino* poder ser analisado como uma forma diminutiva de *pequeno* foi ignorado). Uma vez que este é um adjetivo predicativo, em vez da dependência :MOD, a dependência :ARG0-OF é usada, para representar a frase base equivalente *O sítio é pequenino*. Nesta solução, o LMR não difere substancialmente do AMR padrão.

A frase em PB, por outro lado, está mais próxima da versão EN, apesar de a relativa se encontrar na ordem básica na sequência linear da frase, enquanto em EN esta foi deslocada para o início da frase. Além disso, não há quantificação de grau nesta versão. A notação AMR padrão da estrutura correspondente da frase em EN considera a construção predicativa do adjetivo *small*, tornando-o o nó raiz do gráfico, ignorando o verbo auxiliar (cópula) e ligando-o ao seu sujeito (everything) por uma dependência :DOMAIN. Finalmente, a dependência :LOCATION liga diretamente o adjetivo raiz *small* ao verbo da (chamada) relativa *live-01*.

Ao compararmos a representação em PB, em vez da dependência :DOMAIN padrão do AMR, surpreendentemente encontramos uma :MOD, que não está de acordo com as diretivas do AMR. De resto, as duas representações são basicamente as mesmas.

Na gramática portuguesa, este tipo de estrutura é frequentemente associado às chamadas subordinadas relativas sem antecedente expresso (Velooso, 2013) e pode ser considerado um tipo especial de subordinada interrogativa parcial indireta (Mamede *et al.*, 2012). As diretivas do LMR levam em consideração esta construção complexa, onde o advérbio interrogativo mantém uma dupla relação: (i) a subordinada inteira é um argumento (ou complemento) de um elemento predicativo (ou não) na oração matriz (no caso, o complemento locativo de *tudo é pequeno*); (ii) enquanto o advérbio interrogativo é um argumento (ou complemento) do predicado da oração subordinada (no exemplo, *morar*). Assim, é proposta a seguinte solução alternativa (Figura 14) para a análise LMR da frase em PB:

(PB,72) *Tudo é pequeno onde eu moro.*

Figura 14. Anotação alternativa em LMR para a frase id=72 em PB

```
(ROOT :MAIN (p / pequeno
  :ARG0 (t / tudo)
  :VAUX (e1 / é)
  :LOCATION (o / onde
            :ARGN-OF (m / moro
                      :ARG0 (e2 / eu))))))
```

Repare-se que a solução descritiva do LMR: (i) mantém o adjetivo como o predicado :MAIN mais importante da frase (o nó raiz e as dependências :MAIN são apenas artefatos de notação); (ii) representa explicitamente o verbo auxiliar *ser* (é); (iii) atribui adequadamente ao pronome indefinido *tudo* o seu estatuto de argumento :ARG0, ou seja, o argumento sujeito de *pequeno*; (iv) relaciona adequadamente a oração

principal à subordinada por uma dependência :LOCATION, utilizando o advérbio interrogativo *onde* como o pivô da subordinada (chamada) relativa; (v) expressa o estatuto de argumento :ARGN-OF<sup>13</sup> do advérbio interrogativo *onde* em relação ao verbo da subordinada (*morar*).

Neste sentido, a representação explícita do verbo auxiliar (i) deverá permitir um processamento adicional dos valores temporais-aspetuais que estes elementos veiculam; (ii) o uso de :ARG0 no LMR, em vez da dependência :DOMAIN padrão, simplifica e harmoniza a representação de argumentos verbais e adjetivais, tratando de forma homóloga estas categorias predicativas; (iii) possibilita a expressão da dupla relação semântica observada neste tipo de subordinadas (as chamadas orações relativas sem antecedente exposto), nomeadamente entre o pivô da oração relativa e o verbo principal, por um lado, e entre esse mesmo pivot e o predicado da oração subordinada.

Vejamos, agora, um caso de representação de verbos auxiliares, usando, para tal, a frase com o (id=181). Verifiquemos, em primeiro lugar, que as expressões originais em francês e em inglês são bastante diferentes. Em francês, a frase constrói-se em torno do adjetivo predicativo *indulgents* ‘*indulgentes*’, que analisamos como um predicado com dois argumentos (Baptista, 2005):

FR: *Les enfants doivent être très indulgents envers les grandes personnes.*  
‘As crianças devem ser muito indulgentes com os adultos.’

O complemento *envers les grandes personnes*<sup>14</sup> é analisado como complemento (com estatuto argumental) do adjetivo. Pelo contrário, a frase em inglês estrutura-se a partir do verbo *show* ‘*mostrar*’ sob o qual se encontra o nome predicativo *forbearance* ‘*indulgência*’, analisando-se o complemento *toward grown-up people* ‘*para com as pessoas crescidas*’ como complemento (com estatuto argumental) de *forbearance*. Note-se a presença do advérbio *always* ‘*sempre*’, ausente da versão francesa.

Ora, a representação AMR (Figura 15) não elimina mas substitui o modalizador *should* ‘*dever*’ por um operador específico, *recommend-01* ‘*recomendar*’, de uma forma que parece insuficientemente motivada. Trata ainda a expressão *grown-up people* ‘*pessoas crescidas*’ como complemento de *show* ‘*mostrar*’, apesar de a preposição que introduz esse constituinte ser claramente selecionada pelo nome predicativo e não pelo verbo ditransitivo *show*. Tal solução ignora completamente o carácter predicativo do nome *forbearance* ‘*indulgência*’, bem como o seu domínio argumental, constituído pelos mesmos elementos que são argumentos de *show* ‘*mostrar*’. Note-se ainda que a substituição das formas nominais do texto pelo respetivo lema conduz à perda da informação de plural em *children* ‘*crianças*’. No caso de *grown-up people* ‘*pessoas crescidas*’, o nome *people* é eliminado, o que também faz desaparecer a interpretação *plural* (de coletivo humano) e *genérica*, associada a este nome. Em vez disso, o adjetivo *grown-up* ‘*crescidos*’ figura como argumento de *show* (*mostrar*), em vez de uma análise deste adjetivo como, pelo menos, modificador do nome. Ao não considerar toda a locução como uma unidade lexical perde-se, igualmente, todo o conjunto de conotações que o nome composto expressa.

(EN,181) *Children should always show great forbearance toward grown-up people.*  
‘As crianças devem sempre mostrar grande tolerância para com os adultos.’

<sup>13</sup> No LMR, as dependências :ARGN/:ARGN-OF são também utilizadas para papéis semânticos que não designam o **objeto**, isto é, argumentos locativos, quando estes são selecionados pelo predicado principal (e.g., *viver em Lisboa*).

<sup>14</sup> Interpretamos aqui a locução nominal (ou nome composto) *grandes personnes* no sentido de *pessoas adultas*, numa formulação associada a um discurso infantil, sobretudo considerando a frase correspondente da versão inglesa. Esta interpretação é validada, entre outras fontes, pelo Wiktionary: [https://fr.wiktionary.org/wiki/grande\\_personne](https://fr.wiktionary.org/wiki/grande_personne).

Figura 15. EN; frase id=181

```
(r / recommend-01
  :ARG1 (s / show-01
    :ARG0 (c / child
      :ARG1 (f / forbearance
        :mod (g / great))
      :ARG2 (g2 / grown-up)
      :time (a / always)))
```

No caso da tradução espanhola (Figura 16), feita, como se disse, a partir da versão inglesa, optou-se pela construção adjetival *paciente* (id.) encaixada sob o verbo *mostrarse* ‘mostrar-se’; o nome *mayores*, tradução parcialmente adequada de *grown-up people*, perdendo-se, no entanto, a conotação de adjetivo de relação, presente em expressões como *mis mayores* ‘os meus [parentes] mais velhos’.

(ES,9 [=181]) *Los niños siempre deberían mostrarse muy pacientes con los mayores.*  
 ‘As crianças devem sempre mostrar-se muito pacientes com os adultos.’

Figura 16. ES; frase id=9 [=181]

```
(r / recomendar
  :ARG1 (m / mostrar
    :ARG0 (n / niño
      :ARG1 (p / paciencia
        :grado (m2 / mucho))
      :ARG2 (p2 / persona
        :mod (e / edad
          :grado (m3 / más)))
      :tiempo (s / siempre)))
```

Ora, ainda que a tradução seja adequada, a representação desta frase mostra-se talvez excessivamente ‘colada’ tanto à frase em inglês como à sua representação AMR naquela língua. Desde logo porque o adjetivo *pacientes* é substituído pelo nome predicativo *paciencia* ‘paciência’. Tal acarreta a substituição do advérbio *muy* ‘muito’ por *mucho* ‘muito’, a fim de manter a consistência sintática. Em seguida, porque adota a mesma representação para o auxiliar modal *dever* (id.), substituindo-o pelo operador *recomendar* (id). No mais, as observações feitas a propósito da anotação AMR para a frase inglesa são também relevantes nesta representação.

Ainda assim, note-se que o anotador foi mais consistente no estabelecimento da relação de :GRADO, correspondente à notação inglesa :DEGREE para o modificador do nome predicativo. Note-se, ainda, a conservação do advérbio *siempre* ‘sempre’ e da relação :TIEMPO (~ EN :TIME). O aspeto, quanto a nós mais controverso, em que a anotação se afasta quer do texto quer das diretivas AMR standard consiste no tratamento do nome *mayores*. A representação apresentada corresponderá a expressões como *personas de más edad*, o que, não sendo completamente diferente, não captura grande parte do significado lexicalmente veiculado por um vocábulo como *mayores*.

Observando agora a anotação feita para a tradução da mesma frase em PB (Figura 17), verifica-se que esta frase corresponde palavra por palavra à versão francesa, tendo o anotador sido relativamente consistente com as diretivas AMR standard, por exemplo, no tratamento do modificador *muito* com a relação :DEGREE. Por outro lado, assinala-se a conservação do modal *dever*, embora este seja aparentemente tratado como um verbo pleno. Em contrapartida, o verbo copulativo *ser* não é representado. Ora, à construção adjetival deveria corresponder uma relação de :DOMAIN. Como o verbo *dever* foi analisado como se de um verbo pleno se tratasse e o verbo copulativo se encontra omitido, o adjetivo *indulgente* aparece, assim, como :ARG1 do verbo *dever*. Finalmente, no composto *personas grandes*, o adjetivo é analisado como um simples modificador de *personas*, ignorando a natureza não composicional desta locução (ou nome composto).

(PB,181) *As crianças devem ser muito indulgentes com as pessoas grandes.*

Figura 17. PB; frase id=181

```
(d / dever-01
  :ARG0 (c / criança)
  :ARG1 (i / indulgente
        :degree (m / muito))
  :ARG2 (p / pessoa
        :mod (g / grande)))
```

A tradução portuguesa, tal como a brasileira, segue a versão francesa do texto original. A diferença principal reside em pronominalizar o complemento indireto de *mostrar*, que ocorre num momento anterior do discurso: *As pessoas crescidas são assim*. Na representação correspondente em LMR (Figura 18), e na sequência de Baptista & Mamede (2023), analisamos *mostrar* como um verbo pleno (Baptista & Mamede, 2020a, classe 09I) e o seu complemento direto *uma grande indulgência* como o resultado de uma redução de completiva (no exemplo abaixo, os elementos repetidos e reduzidos encontram-se reconstituídos entre parênteses retos):

*As crianças devem mostrar às pessoas crescidas que [as crianças] têm uma grande indulgência para com [as pessoas crescidas]*

Neste sentido, o verbo auxiliar modal *dever* é explicitamente associado ao verbo principal *mostrar*. A escolha da relação :DEVER resulta, aqui, do contexto específico em que o LMR foi inicialmente desenvolvido: para o processamento de textos jurídicos, a indicação da modalidade deontica é particularmente relevante, pelo que foi desenvolvida esta anotação específica, que se aplica apenas a um pequeno conjunto de verbos auxiliares modais (Baptista & Crismán, 2021) em português. Esta notação pode ser concebida como um tipo específico de :VAUX.

Uma vez que se observa uma correferência entre os argumentos de *mostrar* e os de *indulgência*, cuja ordem parcial é idêntica (*crianças* funciona como :ARG0 e *pessoas grandes* como :ARG1 de ambos os operadores), os argumentos do segundo nome predicativo, apesar de reduzidos, são instanciados por meio das variáveis *c* e *l*. Esta solução de resolução anafórica intrafrásica é inteiramente consistente com as diretivas AMR padrão, apenas diferem no facto de, em LMR, se atribuir o mesmo estatuto ontológico a verbos plenos e a nomes predicativos, o que leva, neste caso, à instanciação dos argumentos do nome predicativo.

(PE,181) *As crianças devem mostrar-lhes uma grande indulgência.*

Figura 18. PE; frase id=181

```
ROOT :MAIN (m / mostrar
  :DEVER (d / devem)
  :ARG0 (c / criança)
  :ARG1 (i / indulgência
        :ARG0 c
        :ARG1 l
        :MOD (g / grande)
  :ARG2 (l / lhes)
```

Para concluir esta análise contrastiva das quatro notações, comentamos a seguinte frase com o id=300:

FR: *J'étais très soucieux car ma panne commençait de m'apparaître comme très grave, et l'eau à boire qui s'épuisait me faisait craindre le pire.*

‘Eu estava muito preocupado porque a minha avaria começava a parecer-me muito grave, e a água potável que se estava a esgotar fazia-me temer o pior.’

Na versão inglesa, esta frase encontra-se cindida em duas (id=299 e id=300), que apresentamos abaixo (Figura 19). Uma vez que as traduções portuguesas seguem a versão francesa (Figuras 21 e 22), numa frase só, compará-las-emos com as devidas adaptações à versão inglesa. No caso da tradução espanhola (Figura 20), que segue fielmente a versão inglesa, apenas está disponível a representação AMR da segunda frase.

(EN,299,300)	<i>I was very much worried , for it was becoming clear to me that the breakdown of plane was extremely serious. And I had so little drinking-water left that I had to fear for the worst.</i>
my fear	‘Eu estava muito preocupado, pois estava a tornar-se claro para mim que a avaria meu avião era extremamente grave. E tinha tão pouca água potável restante que que temer o pior.’
do tive	

Figura 19. EN; frases id=299,300

```
(c2 / cause-01
  :ARG0 (c / clear-06
    :ARG1 (s / serious-02
      :ARG1 (b / break-down-12
        :ARG1 (p / plane
          :poss i))
        :degree (e / extreme))
      :ARG2 (i / i))
    :ARG1 (w / worry-01
      :ARG1 i
      :quant (m / much
        :degree (v / very))))))

(a / and
  :op1 (h3 / have-degree-91
    :ARG1 (w / water
      :purpose (d / drink-01)
      :ARG1-of (l2 / leave-17)
      :ARG1-of (h / have-03
        :ARG0 (i / i)))
    :ARG2 (l / little)
    :ARG3 (s / so)
    :ARG6 (o / obligate-01
      :ARG1 i
      :ARG2 (f / fear-01
        :ARG0 i
        :ARG1 (t / thing
          :ARG1-of (h2 / have-degree-91
            :ARG2 (b / bad-07)
            :ARG3 (m / most)))))))))
```

A primeira observação é a substituição da conjunção subordinada causal *for* pelo construto abstrato *causa-01*. Este construto toma os seguintes argumentos: como :ARG0, a oração subordinada causal (*it was becoming clear to me that...*), e como :ARG1 a oração principal (*I was very much worried*). Em seguida, a construção adjetival de *clear-06* ‘claro’, com a completiva-sujeito deslocada da posição básica, é analisada

como um operador normal, sem recurso à notação standard para as construções adjetivais, com a relação :DOMAIN, e o seu sujeito extraposto como o :ARG1.<sup>15</sup> Tal, ainda que indicado nas diretivas, revela alguma inconsistência da notação. O mesmo sucede com o adjetivo *serious* ‘sério/grave’. Como habitualmente, ambos os verbos copulativos que suportam estes adjetivos são ignorados na notação. Em particular, o valor aspetual inceptivo de *become* ‘tornar-se’ é eliminado.

Além disso, se a construção do adjetivo *clear* poderia ser representada como um predicado com dois argumentos, v.g. *something is clear to someone* (*algo é claro para alguém*), na descrição do predicado verbal *clear-06* ‘clarificar/tornar claro’,<sup>16</sup> considera-se 3 argumentos, um ARG0 com papel de **causa**, um ARG1 com valor de **objeto** e um ARG2 representado a pessoa para quem o ARG1 se torna claro (**experimentador?**). Por outro lado, se a construção de *serious-02* for representada como um predicado com um só argumento, v.g. *something is serious* (*algo é grave*), parece difícil justificar a relação semântica de :ARG1 para o sujeito deste adjetivo (*break-down-12*). Efetivamente, nas *Ontonotes*,<sup>17</sup> este predicado não apresenta sequer um papel ARG0. Vê-se assim, como a associação dos predicados adjetivais a lemas verbais pode não ser inteiramente adequada. A notação destes argumentos como :ARG1 é mais um artefacto do esquema de representação das *Ontonotes* do que uma configuração semântica regular (e generalizável) entre predicados semânticos e os seus argumentos.

Tal perplexidade já não se levanta, por exemplo, no caso do predicado adjetival *worry-01* ‘preocupado’. Efetivamente, a sua estrutura predicativa poderá, eventualmente, ser descrita pela correspondente construção verbal e, tratando-se de um verbo dito ‘psicológico’ (classe 04, Baptista & Mamede, 2020a), corresponder então à estrutura *something cause worries somebody* ‘algo causa preocupa alguém’. Nessa construção, o verbo apresenta um sujeito **causativo** e um complemento **experimentador**, preenchido por um nome humano, aqui o pronome *I* ‘eu’, a que poderia corresponder, pois, a relação de :ARG1. (Os restantes aspetos da representação desta frase já foram analisados anteriormente, pelo que não nos alongaremos mais aqui).

Relativamente à segunda frase, a anotação AMR baseia-se numa conceptualização abstrata de predicados como é caso de *have-degree-91*,<sup>18</sup> que está associado a construções adjetivais que exprimem predicados graduáveis, e.g. , e *have-03*,<sup>19</sup> que corresponde ao verbo *have* ‘ter’ no sentido de posse. Ainda assim, é difícil interpretar, nesta frase, a representação (simplificada aqui):

```
h3 / have-degree-91 :ARG1 (w / water :ARG1-OF (h / have-03 :ARG0 (i / i)))
```

pois, na frase que estamos a analisar, não se verifica tal configuração: encontramos o verbo *have* com um **objeto** *water* ‘água’, quantificado por *so little* ‘tão pouco’. Por outro lado, o segundo verbo *have* (*have-03*) corresponderia à representação do significado associado a “posse”, pelo que a presença dos dois operadores parece, no mínimo, redundante.

Note-se, de seguida, que o auxiliar modal *have* em *I had to fear* ‘eu tinha de temer’ é substituído pelo operador *obligate-01*. Tal ignora a natureza do auxiliar modal, que, por ser transparente às restrições de seleção do verbo auxiliado, *fear* ‘temer’ deveria ter o mesmo sujeito que este verbo. No entanto, este operador aparece com o seu sujeito marcado como um :ARG1, consequência da substituição do auxiliar. Finalmente, a expressão *fear for the worst* ‘temer o pior’ recebe uma representação em que se procurou *decompor* o valor idiomático da construção, em vez de reconhecer a sua não composicionalidade semântica, já lexicalizada,<sup>20</sup> e representá-la sem esta decomposição do significado.

<sup>15</sup> <https://www.isi.edu/~ulf/amr/lib/amr-dict.html#:domain>

<sup>16</sup> <https://propbank.github.io/v3.4.0/frames/clear.html#clear.06> (Repare-se que nenhum exemplo é aqui dado de uma construção verbal com ARG0. Este papel aparece expresso numa construção adjetival e complemento causal introduzido por *from*: *It is clear to me from the analysis* [...] ‘Isso é claro para mim a partir da análise’.)

<sup>17</sup> <https://propbank.github.io/v3.4.0/frames/serious.html#serious.02>

<sup>18</sup> <https://www.isi.edu/~ulf/amr/lib/amr-dict.html#have-degree-91>

<sup>19</sup> <https://propbank.github.io/v3.4.0/frames/have.html#have.03>

<sup>20</sup> <https://dictionary.cambridge.org/dictionary/english/fear-the-worst>

Em relação à frase em espanhol (Figura 20), que corresponde apenas à segunda frase da versão inglesa (id=300), a notação segue de muito perto, como habitualmente, aquela representação em AMR standard. A conjunção *y* ‘e’ é aqui usada com a função de ligar a frase corrente à frase anterior. Contudo, esta conjunção é tratada como outra situação qualquer de coordenação, mas apenas apresentando o operador :OP1, que, em princípio, deveria ligar a conjunção ao primeiro membro da coordenação. No entanto, não há segundo membro da coordenação e respetivo operador :OP2. Assim, a aceitar-se a descrição da função de *y* (como também a de *and*) acima, o primeiro membro da coordenação (OP1) deveria ser a frase anterior. Como o AMR não trata, neste momento, este tipo de relações transféricas (veja-se, no entanto, O’Gorman *et al.* (2018), para uma análise e representação de relações discursivas transféricas), qualquer notação estaria sempre incompleta. Ainda assim, a escolha de uma relação :OP1 parece algo equívoca.

Outro aspeto interessante é a simplificação (e maior proximidade ao texto) da representação do constituinte *tan poca agua potable* ‘tão pouca água potável’, argumento de *quedar* ‘restar’, que se baseia nas palavras do texto e não recorre ao tipo de constructos que se viu em AMR standard. Ainda assim, analisamos esta construção de *quedar* como um predicado com dois argumentos, em que *agua* ‘água’ deveria corresponder ao :ARG0, enquanto o pronome dativo de primeira pessoa *me* corresponderia a um :ARG1.

Finalmente, tal como em inglês, o anotador pretendeu representar, a expressão *lo peor* ‘o pior’, fazendo-o corresponder a elementos que não estão presentes no texto (*malo máximo*).

(ES,15[=300])      *Y me quedaba tan poca agua potable que me temía lo peor.*  
 ‘E tinha tão pouca água potável restante que temia o pior.’

Figura 20. ES; frase id=15[=300]

```
(y2 / y
  :op1 (c / causar
    :ARG0 (q / quedar
      :ARG1 (a / agua
        :mod (p / potable)
        :mod (p2 / poco
          :grado (t / tan)))
      :ARG2 (y / yo))
    :ARG1 (t / temer
      :ARG0 y
      :ARG1 (m / malo
        :grado (m2 / máximo))))))
```

Vejam agora a análise da tradução em português do Brasil, comparando-a com a versão original (francesa). Nesta frase, o tradutor omitiu a oração principal, com o predicado de *soucieux* (*J’étais très soucieux*) e a conjunção causal *car* ‘pois’ que a liga ao resto da frase. Fez igualmente uma profunda transformação da segunda oração coordenada encaixada sob *car*: *et l’eau à boire qui s’épuisait me faisait craindre le pire* é traduzida por *e em, breve já não teria água para beber....* A construção com verbo-operador *faire* ‘fazer’ desaparece, bem como a construção do verbo *s’épuiser* ‘acabar-se/esgotar-se’. Desaparece também a expressão idiomática *craindre le pire* ‘temer o pior’. Estas liberdades do tradutor não permitem, pois, uma comparação direta entre as soluções de anotação adotadas entre as diferentes línguas neste caso, mas apenas um comentário genérico à representação AMR produzida (Figura 21).

(PB,299;300)      *Minha pane começava parecer demasiado grave, e em, <sic> breve já não teria água para beber...*



Figura 21. PB; frase id=299;300

```
(c / começar-01
  :ARG0 (p / pane
    :poss (m / minha))
  :ARG1 (p1 / parecer-01
    :ARG2 (g / grave
      :degree (d / demasiado)))
  :cause (t / ter-01 :polarity -
    :ARG0 (e / eu)
    :ARG1 (a / água)))
```

Começamos por assinalar o tratamento de *começar*, auxiliar de *parecer*, como um verbo pleno, bem como o verbo *parecer*. Ainda que admitindo essa solução, ao arrepio de uma análise mais tradicional de *parecer* como verbo copulativo de uma construção adjetival, não se percebe a relação :ARG2 que liga este verbo ao adjetivo *grave*. Tal como se referiu atrás, o critério de transparência dos auxiliares às restrições de seleção dos respetivos elementos auxiliados (Baptista & Crismán, 2021) deveria apontar para uma análise em que *grave* fosse o elemento predicativo principal desta oração, isto é: (g / grave :ARG0 (p / pane)). O segundo aspeto interessante é o facto de a coordenação com a conjunção *e* ter sido analisada como uma relação **causal**, denotada pelo operador :CAUSE, o que, sem ser implausível, nos parece pouco motivado. Finalmente, repare-se a supressão da locução adverbial temporal *em breve*.

Por último, vejamos a tradução em português europeu e a proposta para a sua anotação em LMR (Figura 22). Esta tradução é bastante mais “fiel” à versão francesa original, apenas tomando a liberdade de modificar *l'eau à boire qui s'épuisait* por *a pouca água que restava para beber*. Tal modificação altera, pois, a dependência do verbo *beber* e insere o quantificador *pouca* associado ao emprego do verbo *restar*.

(PE,300) *Estava bastante inquieto, pois a avaria começava a parecer grave, e a pouca água que restava para beber fazia-me temer o pior.*

Figura 22. PE; frase id=300

```
ROOT :MAIN (i / inquieto
  :ARG0 m
  :VAUX (e / estava)
  :DEGREE (b1 / bastante)
  :CAUSE (p1 / pois
    :OP2 (e / e
      :COORD1 (g / grave
        :ARG0 (a / avaria)
        :VAUX (p2 / parecer
          :VAUX (c / começava
            MWE_CONT (a / a))))
      :COORD2 (f / fazia
        :CAUSE (a / água
          :QUANT (p3 / pouca)
          :ARG0-OF (r / restava
            :ARG0 (q / que)
            :PURPOSE (p4 / para
              :OP2 (b2 / beber)))
          :VOPC (top / temer_o_pior
            :ARG0 (m / me))))))
```

Esta frase permite apresentar vários aspetos interessantes do esquema de anotação em LMR. Desde logo, o uso do operador :OP2, “reutilizado” do AMR standard para ligar as conjunções *pois* e *para* às frases que estas introduzem. Um segundo aspeto é a explicitação das relações de coordenação, através dos operadores

COORD1 e COORD2 e em vez dos operadores genéricos OP1 e OP2 do AMR standard. Aqueles operadores desenvolvem- -se, no entanto, a partir da conjunção coordenativa, tal como em AMR, o que permite manter um estreito paralelismo entre as duas notações.

Analisamos aqui o verbo *parecer*, seguindo uma abordagem bastante tradicional, como um verbo copulativo, que funciona como auxiliar gramatical do adjetivo *grave*. Note-se a recursividade da cadeia de auxiliares *começar a parecer*, bem como a explicitação da preposição que a construção *começar a* + infinitivo determina. Tal permite, entre outras coisas, distinguir diversos valores que este verbo pode apresentar em diferentes perífrases verbais, e.g. *começar por* + infinitivo.

Outro aspeto saliente é a análise das orações relativas. Estas fazem-se ligando o antecedente do pronome relativo ao verbo da relativa por uma relação ARGn-OF ‘invertida’, em que ‘n’ indica a relação semântica que este elemento tem na frase de base da relativa. Finalmente, essa relação é repetida, sem a inversão, entre o verbo da relativa e o pronome relativo.

Por último, introduzimos aqui o conceito de verbo-operador causativo (*Vopc*; Baptista 2005, pp. 202 ss.; M. Gross, 1981). Trata-se de um conjunto de operadores que se aplicam a uma frase, acrescentando-lhe um argumento suplementar, estabelecendo entre a frase de base e esse elemento uma relação de **causa**. No caso, o verbo *fazer* desempenha aqui essa função (simplificando o exemplo): *A água fazia/Vopc # eu temia o pior*. Para este tipo de operadores, o LMR propõe que se definam duas relações: por um lado, :CAUSE, a ligar o verbo-operador ao seu sujeito; por outro lado, a relação :VOPC, a ligar o verbo-operador à frase encaixada. Chama-se, finalmente, a atenção para a análise da expressão idiomática *temer o pior* como um único nó.

Poderia parecer excessivo continuar a fornecer mais exemplos, dado que muitos dos fenómenos abordados anteriormente já foram ilustrados neste texto. Certamente, há nuances e detalhes interessantes que não foram abordados aqui por uma questão de espaço. No entanto, acreditamos que a variedade de situações discutidas oferece uma visão geral, embora já bastante aprofundada, das escolhas e decisões tomadas no desenvolvimento do LMR, destacando as suas diferenças em relação às diretrizes de anotação do AMR e, especialmente, em comparação com práticas de anotação anteriores, em inglês, espanhol e mesmo em português do Brasil. Como é comum em qualquer investigação em curso, há, sem dúvida, muito mais a ser explorado e discutido.

### 3. Conclusão

A partir das observações e das propostas delineadas ao longo deste artigo, destacam-se tanto os desafios enfrentados na implementação das diretivas do AMR standard, quanto as potencialidades introduzidas pela proposta de anotação do LMR. Torna-se claro que, independentemente das possíveis discrepâncias advindas das versões originais ou das escolhas lexicais dos tradutores, existem inconsistências na aplicação das diretivas do AMR, as quais se revelam não somente na anotação do texto em inglês, mas se tornam particularmente pronunciadas na anotação da sua tradução em espanhol, bem como na representação AMR da tradução do texto para o português do Brasil, feita a partir da versão francesa.

A abordagem proposta pelo LMR, que se ancora diretamente no texto, apresenta-se como uma solução interessante e promissora, proporcionando uma representação, quanto a nós, mais adequada, porque mais próxima do texto, e menos sujeita às inconsistências inerentes de um processo de abstração de sentido, que é o objetivo último desta anotação tão complexa. O LMR atenua essas inconsistências, (i) ao descrever de forma equivalente as relações predicado-argumento, sem distinção da categoria morfossintática do predicador; (ii) ao representar de maneira explícita determinadas relações de sentido, ancorando-as diretamente nas palavras da frase que as veiculam, nomeadamente, as expressas por elementos gramaticais como preposições, conjunções e advérbios; (iii) ao construir o grafo que representa o significado da frase a partir de um nó ROOT; (iv) ao permitir que certos modificadores sobre toda a frase possam ficar associados a esse nó ROOT; (v) ao explicitar os valores de operadores gramaticais, tais como os verbos auxiliares, verbos copulativos e verbos-suporte,

reconhecendo a sua contribuição relevante para (parte de) o significado da frase; (vi) ao simplificar a anotação, recorrendo a um pré-processamento para a delimitação e classificação de expressões multipalavra, nomeadamente, de locuções e palavras compostas, bem como de expressões idiomáticas, que passam a ser consideradas como unidades de sentido do texto; (vii) ainda nesse pré-processamento, procede-se à delimitação e classificação de entidades mencionadas (Pessoa, Organização, Locativos, etc.), frequentemente expressões multipalavra, bem como de expressões adverbiais de tempo; (viii) ao limitar ao mínimo a reconstituição de elementos reduzidos, nomeadamente, constituintes anaforicamente reduzidos (a zero), só realizando a resolução de anáfora (a zero ou pronominal) em contexto intrafrásico quando os antecedentes estão explícitos e relegando-se outras formas de anáfora para uma fase de pós-processamento.

Para trabalho futuro, pretende-se alargar o conjunto de textos já anotados em LMR, concluindo a anotação do texto de *O Príncipezinho*, já bastante adiantada, e estendendo-o a textos de diferentes géneros e domínios, ampliando também o conjunto de textos já anotados do domínio jurídico.

Pretende-se desenvolver um conjunto de ferramentas que possibilitem uma implementação mais rápida e eficiente da anotação, nomeadamente: (a) um lematizador que permita associar às formas do texto o respetivo lema e um índice lexical, em particular o identificador unívoco de sentido/construção dos elementos predicativos no léxico-gramática do português; (b) uma ferramenta auxiliar à construção dos grafos LMR, que instancie as posições argumentais dos elementos predicativos do texto, dada a respetiva construção/sentido, permitindo associar-lhes os elementos textuais e assinalar as posições para as quais será necessária a resolução de anáfora; essa ferramenta deverá igualmente validar a consistência formal da representação; (c) uma ferramenta que permita converter grafos em formato gráfico ou no formalismo PENMAN, para facilitar a interpretação da representação por parte dos anotadores; (d) uma ferramenta que permita comparar anotações concorrentes da mesma frase ou expressão, medindo o acordo entre anotadores, e, num segundo momento, medindo o grau de correspondência ou de equivalência entre traduções do mesmo texto em diferentes línguas. Espera-se, também, que um *corpus* mais alargado de textos anotados, incluindo textos de diferentes géneros e domínios, permita, depois, a construção de um parser LMR, que possa produzir automaticamente a partir dos textos uma representação abstrata do respetivo significado, eventualmente sujeita a validação e/ou correção humana subsequente.

#### Agradecimentos / Financiamento

Este trabalho contou com apoio da Fundação para a Ciência e a Tecnologia (FCT) (Proj.Ref. UIDB/50021/2020, DOI: 10.54499/UIDB/50021/2020).

#### Referências

- Anchiêta, Rafael, Marco Cabezudo & Thiago Pardo (2019) SEMA: an extended semantic evaluation metric for AMR. *arXiv e-prints*, pp. arXiv–1905.
- Anchiêta, Rafael (2020) *Abstract meaning representation parsing for the Brazilian Portuguese language*. Tese de doutoramento, Universidade de São Paulo, Brasil.
- Azin, Zahra & Gülşen Eryiğit (2019) Towards Turkish abstract meaning representation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*. Association for Computational Linguistics, pp. 43–47. <https://doi.org/10.18653/v1/P19-2006>
- Banarescu, Laura, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer & Nathan Schneider (2013) Abstract meaning representation for SemBanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pp. 178–186. Disponível em <https://aclanthology.org/W13-2322>
- Baptista, Jorge (2005) *Sintaxe dos predicados nominais com ser de*. Fundação Calouste Gulbenkian/Fundação para a Ciência e a Tecnologia.

- Baptista, Jorge (2010) *Verba dicendi*: A structure looking for verbs. In Takuya; Nakamura, Éric Laporte, Anne Dister & Cédric Fairon (eds.), *Les tables. La grammaire du français par le menu. Mélanges en hommage à Christian Leclère*, 6 in Cahiers du CENTAL. CENTAL/Presses Universitaires de Louvain, Louvain-la-Neuve, pp. 11–20.
- Baptista, Jorge (2012) ViPEr: a lexicon-grammar of European Portuguese verbs. In *Actes du 31e Colloque International sur le Lexique et la Grammaire*. Nové Hradý, pp. 10–17.
- Baptista, Jorge (2013) ViPEr: Uma base de dados de construções léxico-sintáticas de verbos do Português Europeu. In *Actas do XXVIII Encontro da APL – Textos Seleccionados*. APL & Colibri, pp. 111–129.
- Baptista, Jorge (2024a) *Lexicalized meaning representation – guidelines*. Disponível em [https://gitlab.hlt.inesc-id.pt/u000803/lmr4pt/-/blob/master/LMR\\_Directives\\_20240501.pdf](https://gitlab.hlt.inesc-id.pt/u000803/lmr4pt/-/blob/master/LMR_Directives_20240501.pdf)
- Baptista, Jorge (2024b) LMR4PT – *O Príncipezinho* (annotated version 01). Disponível em [https://gitlab.hlt.inesc-id.pt/u000803/lmr4pt/-/blob/master/public/LMR4PT\\_Principezinho.pdf](https://gitlab.hlt.inesc-id.pt/u000803/lmr4pt/-/blob/master/public/LMR4PT_Principezinho.pdf)
- Baptista, Jorge & Nuno Mamede (2020a) *Dicionário gramatical de verbos do Português Europeu*. Universidade do Algarve Editora.
- Baptista, Jorge & Nuno Mamede (2020b) Syntactic transformations in rule-based parsing of support verb constructions: examples from European Portuguese. In Alberto Simões, Pedro Rangel Henriques, & Ricardo Queirós (eds.), *9th Symposium on Languages, Applications and Technologies* (SLATE 2020). Schloss Dagstuhl–Leibniz-Zentrum für Informatik, pp. 11:1–11:14. <https://doi.org/10.4230/OASICS.SLATE.2020.11>
- Baptista, Jorge & Nuno Mamede (2023) Verbo-suporte escondido “com rabo de fora”: Construções complexas com verbo-suporte reduzido. *Revista da Associação Portuguesa de Linguística*, (10), pp. 41–57. <https://doi.org/10.26334/2183-9077/rapln10ano2023a3>
- Baptista, Jorge, Nuno Mamede & Sónia Reis (2022) Support verb constructions across the ocean sea. In *18th Workshop on Multiword Expressions @ LREC2022*. ACL, pp. 26–36.
- Baptista, Jorge & Rafael Crismán (2021) Auxiliary verb constructions in Portuguese and Spanish. A comparative study Construcciones verbales auxiliares en portugués y español. Un estudio comparativo. *Revista de Linguas Modernas* 34, pp. 39–57. <https://doi.org/10.15517/rlm.v0i34.41462>
- Baptista, Jorge, Simone Pereira & Nuno Mamede (2016) ZAC: Zero Anaphora Corpus. In *Proceedings: Workshop on Corpora and Tools for Processing Corpora* (colocated with PROPOR 2016), pp. 38–45.
- Bonial, Claire, Bianca Badarau, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Tim O’Gorman, Martha Palmer & Nathan Schneider (2018) Abstract meaning representation of constructions: the more we include, the better the representation. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation* (LREC 2018). European Language Resources Association (ELRA), pp. 1677–1684. Disponível em <https://aclanthology.org/L18-1266>
- Català, Dolors, Jorge Baptista & Cristina Palma (2020) Problèmes formels concernant la traduction des adverbes composés (espagnol/portugais). *Langues & Parole* 5, pp. 67–82. <https://doi.org/10.5565/rev/languesparole.64>
- Damonte, Marco, Shay Cohen & Giorgio Satta (2017) An Incremental Parser for Abstract Meaning Representation. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics* (Vol. 1). European Language Resources Association (ELRA), pp. 536–546. Disponível em <https://aclanthology.org/E17-1051>
- Damonte, Marco & Shay Cohen (2018) Cross-lingual abstract meaning representation parsing. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Vol. 1). Association for Computational Linguistics, pp. 1146–1155. Disponível em <https://aclanthology.org/N18-1104>
- Duran, Magali & Sandra Aluísio (2015) Automatic generation of a lexical resource to support semantic role labeling in Portuguese. In *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics*. Association for Computational Linguistics, pp. 216–221. <https://doi.org/10.18653/v1/S15-1026>

- Galvão, Ana, Jorge Baptista & Nuno Mamede (2019a) New developments on processing European Portuguese verbal idioms. *Proceedings: STIL 2019 – Symposium in Information and Human Language Technology*. Sociedade Brasileira de Computação, pp. 229–238.
- Galvão, Ana, Jorge Baptista & Nuno Mamede (2019b) Processing European Portuguese verbal idioms: from the lexicon-grammar to a rule-based parser. In Glória Corpas Pastor, Ruslan Mitko, Kunilovskaya, M., and Losey León, M. A. (eds.), *Computational and Corpus-based Phraseology. Proceedings of the Third International Conference EUROPHRAS 2019, Malaga, Spain, September 25–27, 2019*. Tradulex, pp. 70–77. [https://doi.org/10.26615/978-2-9701095-6-3\\_009](https://doi.org/10.26615/978-2-9701095-6-3_009)
- Gross, Maurice (1981) Les bases empiriques de la notion de prédicat sémantique, *Langages* 63. Larousse, pp. 7–52.
- Gross, Maurice (1996) Lexicon-Grammar. In Keith Brown & Jim Miller (eds.), *Concise encyclopedia of syntactic theories*. Pergamon, pp. 244–259.
- Gross, Maurice (1998) La fonction sémantique des verbes supports. *Travaux de linguistique : revue internationale de linguistique française* 37 (1), pp. 25–46.
- Harris, Zellig (1991) *A theory of language and information: a mathematical approach*. Clarendon Press.
- Hovy, Eduard & Julia Lavid (2010) Towards a ‘science’ of corpus annotation: a new methodological challenge for corpus linguistics. *International Journal of Translation* 22 (1), pp. 13–36.
- Leclère, Christian (1976) Datifs syntaxiques et datif éthique. In Jean-Claude Chevalier & Maurice Gross (eds.), *Méthodes en grammaire française*, Klincksieck, pp. 73–96.
- Li, Bin, Yuan Wen, Weiguang Qu, Lijun Bu & Nianwen Xue (2016) Annotating the *Little Prince* with Chinese AMRs. In *Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016 (LAW-X 2016)*. Association for Computational Linguistics, pp. 7–15. <https://doi.org/10.18653/v1/W16-1702>
- Linh, Ha, & Huyen Nguyen (2019) A case study on meaning representation for Vietnamese. In *Proceedings of the First International Workshop on Designing Meaning Representations*. Association for Computational Linguistics, pp. 148–153. <https://doi.org/10.18653/v1/W19-3317>
- Mamede, Nuno, Jorge Baptista, Cláudio Diniz & Vera Cabarrão (2012) STRING: A hybrid statistical and rule-based natural language processing chain for Portuguese. In *Proceedings: Computational Processing of the Portuguese Language. 10th International Conference on Computational Processing of Portuguese (PROPOR 2012), Coimbra, Portugal, April 17–20, 2012*. Disponível em <http://www.inesc-id.pt/ficheiros/publicacoes/8578.pdf>
- Mamede, Nuno, & Jorge Baptista (2016) Nomenclature of chunks and dependencies in Portuguese XIP Grammar 4.5 [Relatório Técnico]. L2F-Spoken Language Laboratory / INESC-ID.
- Matthiessen Christian & John Bateman (1991) Text generation and systemic-functional linguistics: experiences from English and Japanese. Pinter Publishers.
- Miguel Abreira, Noelia (2017) *A study towards Spanish abstract meaning representation*. Dissertação de mestrado, Universidade del País Vasco, Espanha.
- Molinier, Christian, & Françoise Levrier (2000) *Grammaire des adverbes : description des formes en-ment* (Vol. 33). Librairie Droz.
- Müller, Izabela, Nuno Mamede & Jorge Baptista (2023) Advérbios compostos do Português do Brasil. *Revista da Associação Portuguesa de Linguística*, (10), pp. 230–250. <https://doi.org/10.26334/2183-9077/rapln10ano2023a13>
- O’Gorman, Tim, Michael Regan, Kira Griffitt, Ulf Hermjakob, Kevin Knight & Martha Palmer (2018) AMR Beyond the sentence: the multi-sentence AMR corpus. In *Proceedings of the 27th International Conference on Computational Linguistics*. Association for Computational Linguistics, pp. 3693–3702. Disponível em <https://aclanthology.org/C18-1313>
- Oral, Elif, Ali Acar & Gülşen Eryiğit (2022) Abstract meaning representation of Turkish. *Natural Language Engineering* 30 (1), pp. 1–30. <https://doi.org/10.1017/S1351324922000183>

- Palma, Cristina (2009) *Expressões fixas adverbiais: descrição léxico-sintáctica e subsídios para um estudo contrastivo português-espanhol*. Dissertação de mestrado. Universidade do Algarve, Portugal.
- Pustejovsky, James, Ken Lai & Nianwen Xue (2019) Modeling quantification and scope in abstract meaning representations. In *Proceedings of the First International Workshop on Designing Meaning Representations*. Association for Computational Linguistics, pp. 28–33. <https://doi.org/10.18653/v1/W19-3303>
- Seno, Eloize, Helena Caseli, Marcio Inácio, Rafael Anchieta & Renata Ramisch (2022) XPTA: um parser AMR para o Português baseado em uma abordagem entre línguas. *Linguística*, 14 (1), pp. 49–68. <https://doi.org/10.21814/lm.14.1.359>
- Takhshid, Reza, Razie-Sadat Shojaei, Zahra Azin & Mohammad Bahrani (2022) Persian abstract meaning representation. *ArXiv*. <https://doi.org/10.48550/arXiv.2205.07712>
- Van Peteghem, Marleen (2006) Le datif en français : un cas structural. *Journal of French Language Studies* 16 (1), pp. 93–110. <https://doi.org/10.1017/S0959269506002286>
- Veloso, Rita (2013) Subordinação relativa. In Raposo, E. B. P. *et al.* (orgs.). *Gramática do português*. Fundação Calouste Gulbenkian, pp. 2059–2134.
- Wein, Shira, & Julia Bonn (2023) Comparing UMR and cross-lingual adaptations of AMR. In *Proceedings of the Fourth International Workshop on Designing Meaning Representations*. Association for Computational Linguistics, pp. 23–33. Disponível em <https://aclanthology.org/2023.dmr-1.3>
- Weischedel, Ralph, Martha Palmer, Mitchell Marcus, Eduard Hovy, Sameer Pradhan, Lance Ramshaw, Nianwen Xue, Ann Taylor, Jeff Kaufman, Michelle Franchini, Mohammed El-Bachouti, Robert Belvin & Ann Houston (2013) *OntoNotes Release 5.0 (LDC2013T19)*. Web Download. Linguistic Data Consortium. <https://doi.org/10.35111/xmhb-2b84>
- Xue, Nianwen, Ondřej Bojar, Jan Hajič, Martha Palmer, Zdeňka Urešová & Xiuhong Zhang (2014) Not an interlingua, but close: comparison of English AMRs to Chinese and Czech. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. European Language Resources Association (ELRA), pp. 1765–1772. Disponível em [http://www.lrec-conf.org/proceedings/lrec2014/pdf/384\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/384_Paper.pdf)